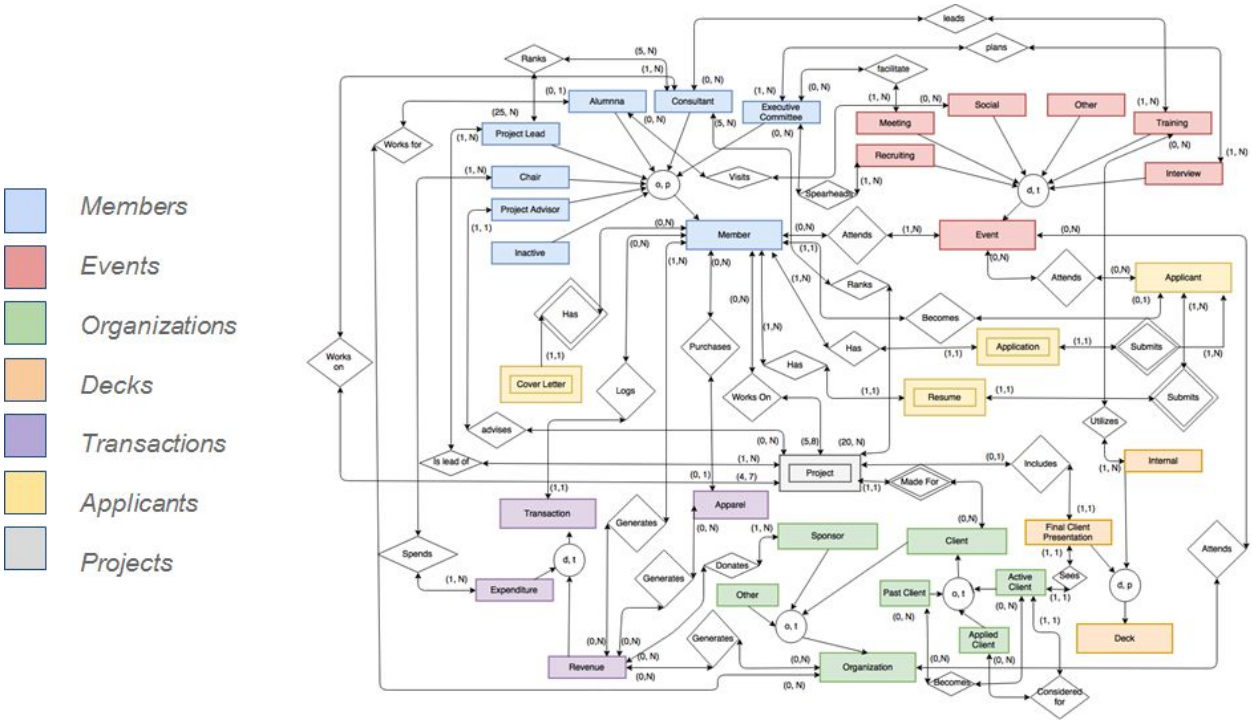
Final Report: The Berkeley Group - Team 3

Taylor Lyberger, Justin Rezende, Lauren Hanlon, Nicole Benun,
Joao Drummond, Achilleas Ghinis, Emily Rah, Matt Campbell

Executive Summary

Client Description: The Berkeley Group (TBG) is a student run consulting firm that provides management consulting services to Bay Area nonprofits. TBG was founded in 2003 by four UC Berkeley students, and it has assisted over 100 social sector organizations in the San Francisco Bay Area. TBG believes that helping social sector organizations with their business challenges is a unique way to create sustainable change in its community, and the organization strives to provide the highest quality services to its clients in order to maximize their capacity for social impact.

Simplified EER Diagram:



Relational Schema:

MEMBERS

1. **Member**(Member_ID, M_Fname, M_Lname, M_Email, M_Phone, M_Address, Graduation_year, Semester_joined, Applicant_ID[6], Gender, Status)
   a. **Executive_committee**(Executive_ID[1], Semester, Position)
      i. **Exec_Position**(Position, Position_description)
   b. **Consultant**(Consultant_ID[1],Project_ID[12])
   c. **Alumni**(Alumni_ID[1], Start_date, End_date, Grad_School_Name, Organization_name[3], Occupation, New_phone, New_email)
   d. **Project_lead**(Project_Lead_ID[1], Project_ID[12])
   e. **Chair**(Chair_ID[1], Semester, Position)
      i. **Chair_Position**(Position, Position_description)
   f. **Project_Advisor**(Project_Advisor_ID[1], Project_ID[12])
   g. **Inactive**(Inactive_ID[1], Semester_inactive, Number_Semesters_Inactive, Number_Semesters_Active)

EVENTS

2. **Events**(Event_ID, Organizer[1], E_Date, E_Time, E_Location, ASUC_Sponsored, Est_Attendance, Budget)
   a. **Recruiting**(Event_ID[2], Lead_Member_ID[1a])
   b. **Interview**(Event_ID[2], Applicant_ID[6], Interview_ID[1a])
   c. **Meeting**(Event_ID[2], Agenda)
   d. **Social**(Event_ID[2], Transportation_type)
   e. **Training**(Event_ID[2], Training_name, Presenter[1])
   f. **Other**(Event_ID[2], Other_name)

DECKS

4. **Deck**(Deck_ID, Title, Date)
   a. **Internal**(Deck_ID[4], Type, Event_Created_For[2])
   b. **Final_client_presentation**(Deck_ID[4], Project_ID[12], Organization_ID[3], Location, Feedback)

ORGANIZATIONS

3. **Organization**(Organization_ID, O_Name, O_Address, O_Phone, O_Email)
   a. **Client**(Client_Organization_ID[3], Source_Ref)
      i. **Active_client**(Act_Organization_ID[3], Active_Date)
      ii. **Applied_client**(App_Organization_ID[3], Application_date, Application_status)
      iii. **Past_client**(Organization_ID[3], Date_Last_Active, Growth_etric)
   b. **Sponsor**(Organization_ID[3], Type)
   c. **Org_Other**(Organization_ID[3], Type)
   d. **Org_Industry**(Organization_ID[3],Industry)

TRANSACTIONS

5. **Transaction**(Transaction_ID, Amount, Member_ID[1], Purpose, Date_Logged)
   a. **Expenditure**(Transaction_ID[5], Date_Spent, Date_Reimbursed)
   b. **Revenue**(Transaction_ID[5], Date_Given)

APPAREL

6. **Apparel**(Apparel_ID, Type, Order_Cost, Sale_Price, Manufacturer, Purchaser_ID[1])

APPLICANTS

7. **Applicant**(Applicant_ID, A_fname, A_lname, A_Email, A_Phone, A_Address, Graduation_year, Major(s), GPA, No_Apps, Status, Teamwork_Score, Presentation_Score, Quant_Score, Nonprofit_Score, Compatibility_Score)
8. **Application**(Application_ID, Applicant_ID[6], File_type, Application_attachment, Source_Ref)
9. **Resume**(Resume_ID, Applicant_ID[6], File_type, Resume_attachment)
10. **Cover_letter**(CL_ID, Applicant_ID[6], File_type, CL_attachment)
11. **Transcript**(Transcript_ID, Applicant_ID[6], File_type, Transcript_attachment)

PROJECTS

12. **Project**(Project_ID, Semester, P_Name, Organization_ID[3], Project_Lead_ID[1D], Final_Deck[4])

MULTIVALUED ATTRIBUTES

31. **M_Majors**(Member_ID[1], M_Major)
32. **M_Race**(Member_ID[1], M_Race)
33. **Authors**(Deck_ID[4], Member_ID[1])
34. **A_Majors**(Applicant_ID[6], Major)
35. **Advertising_Avenues**(Event_ID[2], Avenue)
36. **Project_Members**(Project_ID[12], Member_ID[1])
37. **Org_Comments**(Organization_ID[3], Comment, Comment_Date)
38. **Project_Comments**(Project_ID[12], Comment, Comment_Date)
39. **Client_App_Reviewers**(App_Organization_ID[3Ai], Member_ID[1])
40. **Alumni_Invited**(Social_ID[2D], Alumni_ID[1c])

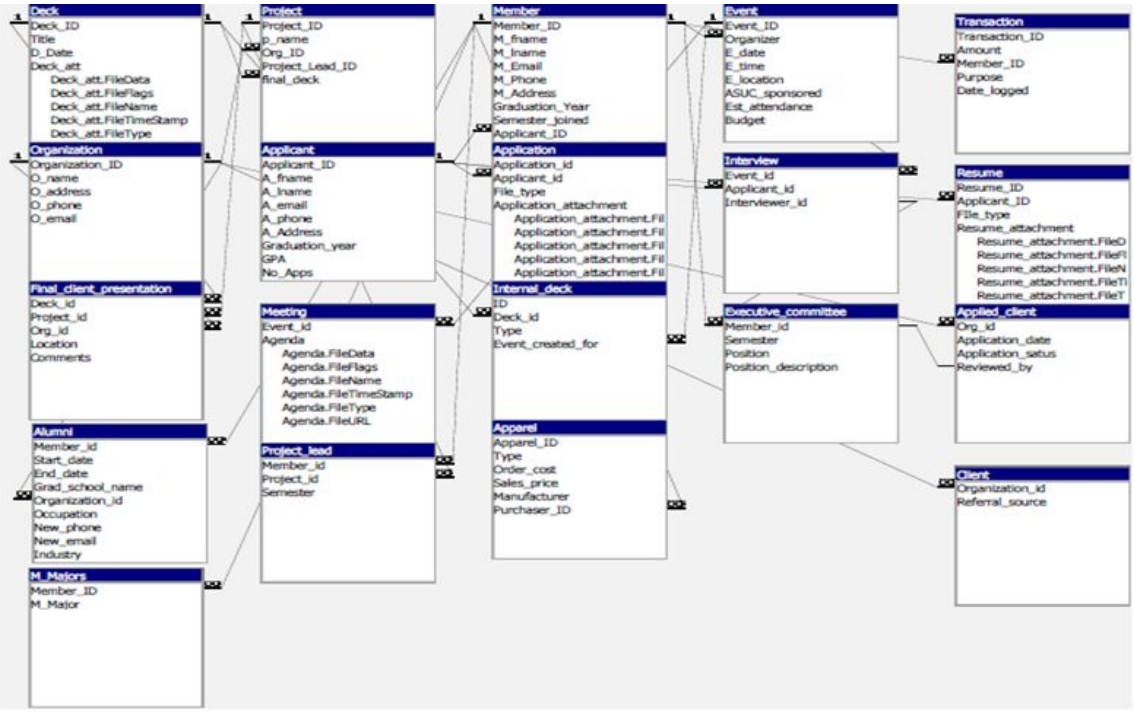
RELATIONSHIPS

13. **Leads**(Consultant_ID[1B], Training_ID[2E], Training_Deck)
14. **Plans**(Executive_ID[1A], Interview_ID[2B], Scheduling_System)
15. **Facilitates**(Executive_ID[1A], Meeting_ID[2C], Planning_Resource, Meeting_Deck)
16. **Spearheads**(Executive_ID[1A], Recruiting_ID[2A], Recruiting_Role, Hours_Contributed)
17. **Member_Attends**(Member_ID[1], Event_ID[2], Clock-In, Clock-Out)
18. **Applicant_Attends**(Applicant_ID[6], Event_ID[2], Referral_Source)
19. **Works_On**(Consultant_ID[1], Project_ID[12], Start_Date, End_Date)
20. **Utilizes**(Training_ID[2E], Int_Deck_ID[4A], Creation_Date, Update_Date)
21. **Made_For**(Project_ID[12], Client_ID[3A], Project_Scope, Semester, Contact)
22. **Spends**(Chair_ID[1E], Expenditure_ID[5A], Item, Price, Quantity, Money_Source)
23. **Member_Generates**(Member_ID[1], Revenue_ID[5B], Rev_Type)
24. **Donates**(Sponsor_ID[3B], Revenue_ID[5B], Rev_type)
25. **Org_Attends**(Organization_ID[3], Event_ID[2], Representative)
26. **App_Generates**(Revenue_ID[5B], Apparel_ID[11])
27. **Org_Generates**(Revenue_ID[5B], Organization_ID[3], Rev_Type)
28. **Becomes**(App_Organization_ID[3Aii], Act_Organization_ID[3Ai], Sem_Accepted)
29. **Consultant_Rank**(Project_Lead_ID[1D], Consultant_ID[1B], Semester, Rank)
30. **Project_Rank**(Consultant_ID[1B], Project_ID[12], Rank)

SEMESTER

41. **Date_Translate**(Date, Semester)

Simplified Access Relational Design:



Queries:

Query 1: Applicant Score

The goal of this query is to return a table of all current applicants and their information including scores, GPA, and major. Fake applicant data with category scores and holistic candidate scores for theoretical applicants could be collected from TBG. Using this data a multilinear regression model could identify the relative importance of each category, and using these values future applicants can be scored and ranked to organize thinking around application process.

```
SELECT Applicant.A_fname, Applicant.A_lname, ApplicantGPA,
Applicant.Teamwork_Score, Applicant.Presentation_Score, Applicant.Quant_Score,
Applicant.Nonprofit_Score, Applicant.Compatibility_Score, A_Majors.Major
FROM Applicant INNER JOIN A_Majors ON Applicant.[Applicant_ID] =
A_Majors.[Applicant_ID];
```

Query 2: Marketing Strategy Success

This query pulls data from applications regarding how an applicant heard about TBG. This will then be exported and analyzed to see which sources are increasing gross applications, which are most reported by applicants, and which have the highest “return on investment” in terms of

work hours spent vs. applicant turnout. Additionally a regression analysis can be performed on trends in this ROI (Applicants/Hour of Effort) to forecast the most cost-effective strategy for the upcoming semester.

```
SELECT Application.[Referral_Source], Application.[Semester],
Count(Application.Referral_Source) AS CountOfReferral_Source
FROM Application
GROUP BY Application.[Referral_Source], Application.[Semester];
```

Query 3: Alumni Info by Industry and/or Company

This query is designed to pull all alumni at a certain company or at a more high level in an industry based on user input. The returned alumni will be ranked by number of events they have attended with the user, number of project worked on together, and by the closest year of graduation. The data will be processed using cluster analysis and the output of the analysis will show the member's closeness of relationships with alumni through a dendrogram. This way current members are put in touch with alumni still interested and involved with the club as well as ones they are more likely to know.

In the userform, the user puts their name (Justin Rezende) and industry or company (Bain) they are looking for.

```
@M_fname = 'Justin'
@M_lname = 'Rezende'
@Organization_name = 'Bain'
SQL>  SELECT Grad_school_name, Organization_name, Occpation, Industry,
New_email Member_ID AS 'Alumni_ID'
FROM Alumni
WHERE Member_ID IN (SELECT Member_ID FROM Alumni WHERE Organization_name =
@Organization_name FROM Alumni);
SQL >  CREATE TABLE Events_attended AS
SELECT a.Member_ID, DISTINCT a.Event_ID
FROM Attends a, Member m
WHERE a.Member_ID == (SELECT Member_ID FROM Member WHERE M_fname == @M_fname
and M_lname == @M_lname FROM Member);
SQL > SELECT a.Member_ID
FROM Alumni a
WHERE a.Member_ID IN (SELECT Member_ID FROM Events_attended);
```

Query 4: Project Matching

Consultants rank projects and project leaders rank consultants. These preferences are pulled and exported to R to run a stable marriage algorithm resulting in a suggested project pairing.

In the userform the user provides the current semester.

```
SELECT Consultant_Rank.Project_lead, Consultant_Rank.Consultant,
Consultant_Rank.Rank
FROM Consultant_Rank;
SELECT Project_Rank.Consultant, Project_Rank.Project_lead, Project_Rank.Rank
FROM Project_Rank;
```

Query 5: Budget calculation & forecasting

This query will select a month to month budget estimate from the most recent 3 years. This will be used in a budget calculation and forecasting model, using Holt-Winter's method to account for year-over-year cost increases, semester variations, and monthly variations.

```
SELECT Transaction.[Amount], Month(Transaction.[Date_logged]),
Year(Transaction.[Date_logged])
FROM Transaction
WHERE Year(Transaction.[Date_logged]) > Year(Transaction.[Date_logged]) - 4;
```

Client Introduction and Expansion

Client: The Berkeley Group (TBG) is a student run consulting firm that provides management consulting services to Bay Area nonprofits.

Client Contact: Adithya Iyengar is the current technology chair at TBG, so he has access to and knowledge of all the TBG data. He currently maintains all TBG websites, where much of the group's data is stored. Adithya and team-member Justin worked on a project together last semester.
Email: adithya.iyengar@berkeley.edu. *Phone*: 919-601-2582.

Data: TBG currently stores various data sets in many Excel files, Google sheets, and websites. The data they currently collect and store includes information on both current members and alumni, such as internships, major, and contact info. Trying to contact an alumnus by tracking down this information can be very difficult if it is not centralized and many platforms must be searched. TBG retains information from past projects, including recommendations made to the client, and also keeps information on nonprofit organizations who have previously applied to be client partners. This client information is especially messy, and could be better leveraged to track potential future clients in a simpler database.

Additionally, there is certain data that TBG does not currently store that could be beneficial to their operations. An example of this is the contact information for TBG applicants. When applying to TBG, students are asked for contact and other basic information, including graduation date. Functionally storing data collected from the applicant portal would allow TBG to filter candidates based on GPA and/or graduation date in order to reach out to previous applicants and encourage them to apply for the next recruitment cycle. Consolidating all of the various data collected by TBG would make it much easier for all members to access all parts of it,

as finding information amidst a mess of Google Drive, Dropbox, and website links is very challenging.

Further conversation with Adithya revealed that all TBG member data is still stored on spreadsheets, which are linked through our internal website. These have to be updated manually and are often out of date with regards to current alumni employment, contact information, and involvement/role in the club. There are about 200 total consultants on file, with about 20 new consultants added each year. Previous applicant data which is collected but not utilized at the moment includes information on previous applicants. There are thousands of students who have expressed interest in TBG over multiple semesters, and each student record has an associated year and major. Unfortunately, this data is not maintained somewhere centralized, so cannot be leveraged to reaching out to people during recruitment. The data currently sits in a Google Sheet, which grows but doesn't seem to have a purpose at the moment, due to its ineffective setup.

Other miscellaneous data collected by TBG includes resumes and cover letters from past members. Linking these types of files to the member's current jobs could be helpful for current members who are shaping applications for those specific firms. Additionally, TBG has a large database of past projects in Dropbox that could be compiled and sorted in a more accessible way to reference easily. Finally, TBG has data kept on nearly one hundred past clients, including contact information, main point of contact, summaries of work done, and the client's expressed desire to work with TBG again. The data on past clients is also accompanied by data on hundreds of potential clients. This potential client data is not maintained very well, and increasing the efficiency in flagging nearly selected applicants could help target nonprofits for future partnerships.

Availability of MS Access: All team members, including Adithya, are Berkeley students, so TBG has access to MS Access.

Number of Employees: About 40 active TBG members, along with hundreds of alumni, who retain access to TBG materials and regularly update their contact and employment info.

Enthusiasm: 8. Adithya would love our help with cleaning up and centralizing TBG's data resources, one of his current high priority projects. The only concern he has is the privacy of the data because TBG signs very strict nondisclosure agreements with their partner nonprofits, but he was greatly assuaged by the assurance that our database would not require real data to set up; only an understanding of the types of records necessary.

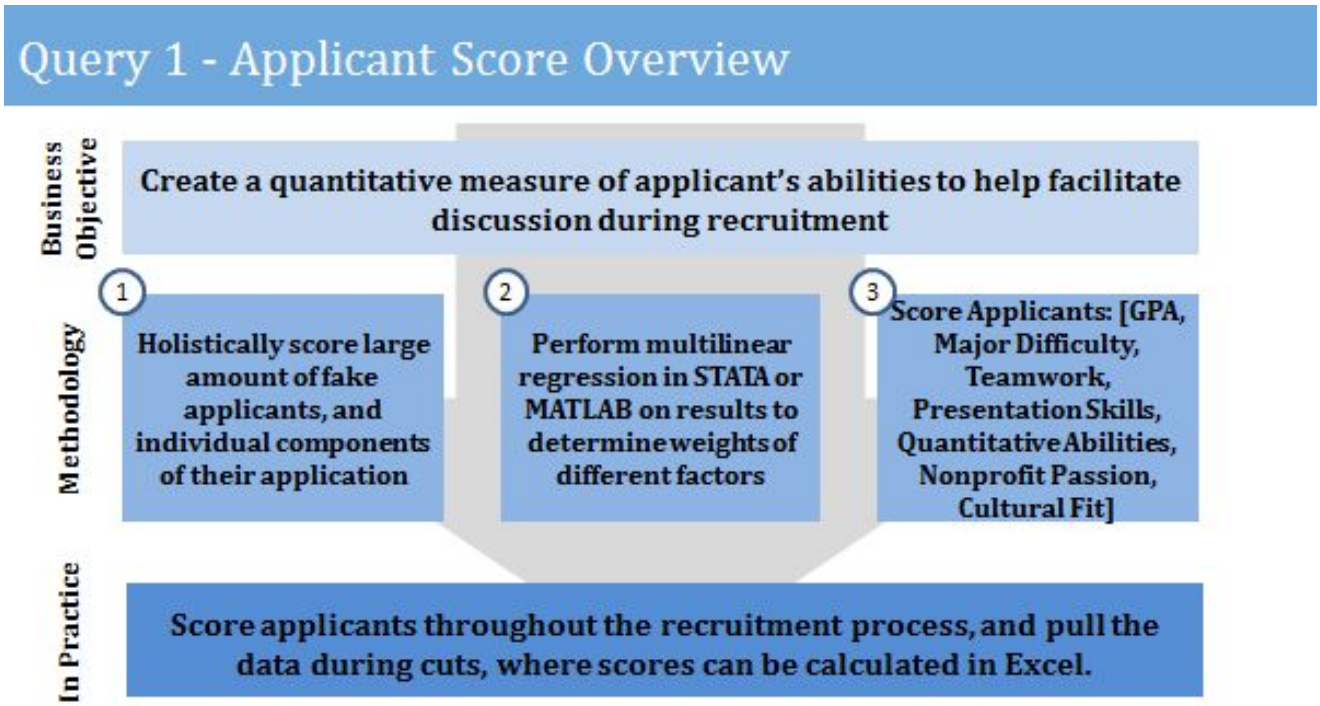
Proposed Benefits: A more organized database would aid in maintaining records to expand the group's reach both on campus and in the community. The ability to update a TBG member's job, have it reflected in all other forms, and shifting his or her resume and cover letters into folders linked to their company would be extremely useful to other current members or alumni looking into that company. Keeping information of past student applicants updated and organized would help with recruitment. Maintaining a working database with past and potential clients could lead to more targeted client outreach for projects believed to be the best fit for TBG. TBG also struggles to manage its budget failing to account for increases in cost over time and semesterly

trends, and would like help with this. TBG can at time work hard not smart, and are hoping to streamline several time consuming practices including application reviews, project team formations, and new member outreach.

Queries Expanded

SQL for all the Queries is in the Executive Summary. In this section we will share more about the need for the queries, the implementation, and the feedback from the client.

Query 1

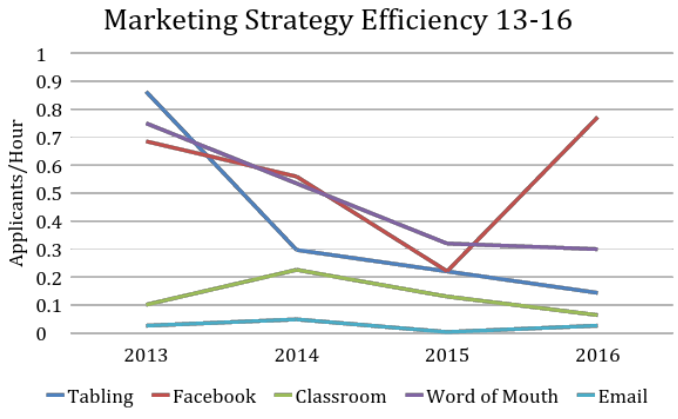


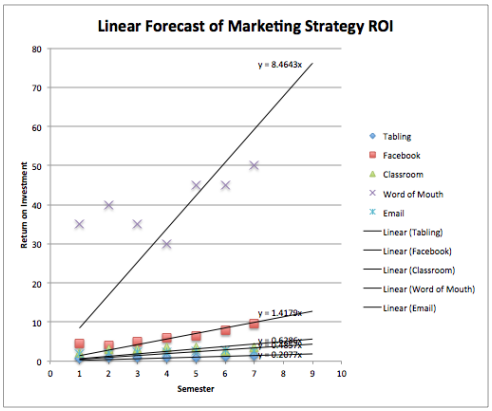
The collected data for training would be a 1xN holistic score vector (Y), and a 7xN individual category score matrix (X). From this the multilinear regression solving:

$$y_i = \beta_0 + \beta_{1}x_{i,1} + \beta_{2}x_{i,2} + \ldots + \beta_{p}x_{i,p}$$

or for all the applicants in matrix notation $Y = X\beta$ for β would find the relative weights of each category. New applicants category scores can the be converted to holistic scores to give an objective measure to assist in a process that tends to stall with many debates over the relative strength of candidates. Client feedback for this Query was not very positive, as TBG fears it would lead to getting new consultants, who were too similar preventing the club from diversifying its thought base, and decreasing TBG's ability to add value to their clients.

Query 2

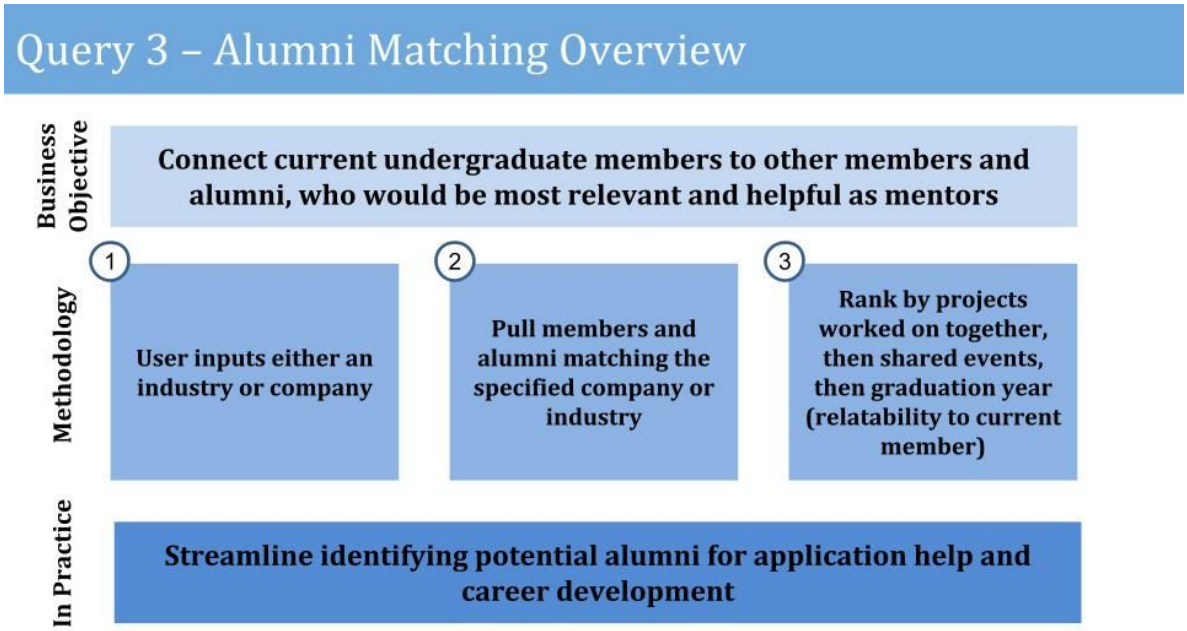




The data collected from this query would be the number of applicants per cycle who cited a marketing source as how they heard about TBG. The ROI is calculated by the number of applicants who cited that source divided by the number of hours spent by TBG on that marketing source. Approximate number of hours per semester are as follows: Tabling ≈ 120, Facebook ≈ 10, Classroom ≈ 10, Email ≈ 20 and Word of Mouth ≈ 1. We then built multiple linear regression models per marketing strategy, to predict ROI in upcoming semesters. The general equation for

each linear regression model is $Y = \beta_0 + X\beta$. We decided to set the y-intercept $\beta_0 = 0$ then found the slopes (β's) to complete each equation. The equations are as follows (per marketing strategy): Tabling = 0.2077*x*, Facebook = 1.4179*x*, Classroom = 0.6286*x*, Email = 0.4857*x*, WordofMouth = 8.4643*x*. Client feedback for Query 2 was very positive as the leadership worries each semester that they uselessly expand hours spent on marketing. Given this query's results, marketing efforts can be more targeted and focus on strategies that give the highest ROI.

Query 3



In order to execute the analysis for this query, we use a hierarchical cluster analysis. Each member represents an element in the cluster analysis, and their proximity to other members and alumni is determined by the number of projects they have worked on together, the number of events that they have both attended, and the difference in number of years between graduation. Because we want a single numerical value assigned to the distance between two members, we will use the following model to determine this value:

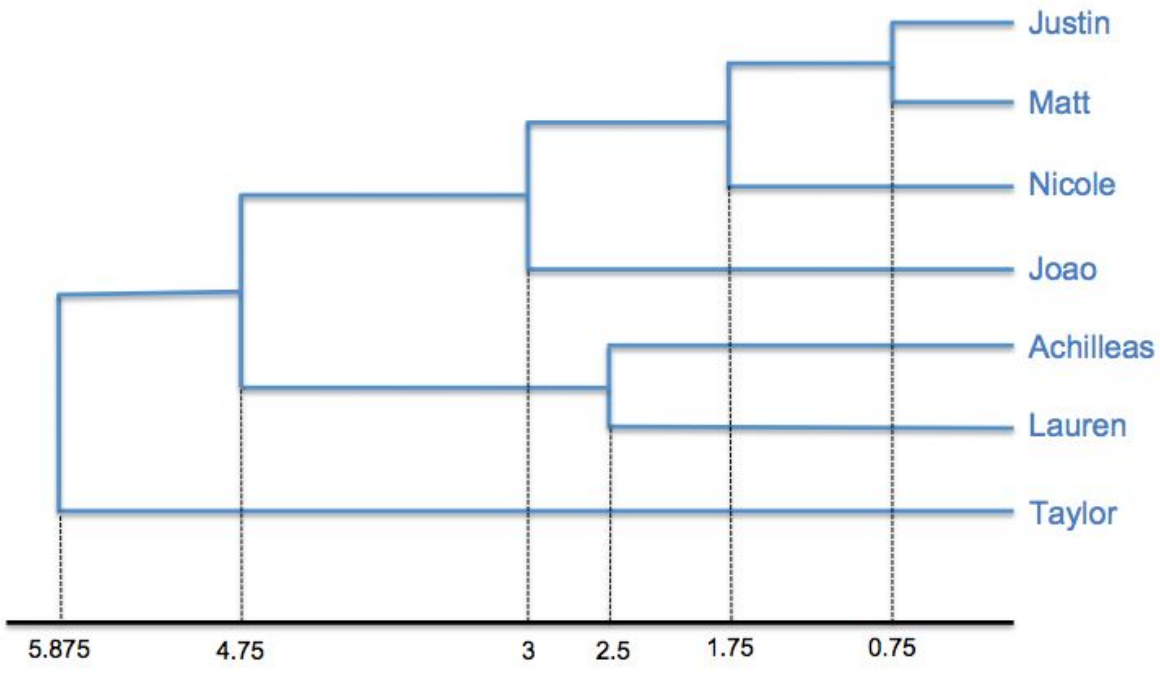
Value Assignment	**0**	**1**	**2**	**3**	**4**	**5**
Number of similar projects	5 or more	4	3	2	1	0
Number of similar events	12 or more	8-10	5-7	3-4	1-2	0
Number of years apart	0	1	2	3	4	5 or more

Sum the value assignments to retrieve the "distance" between two members

Sample Distance Assignments:

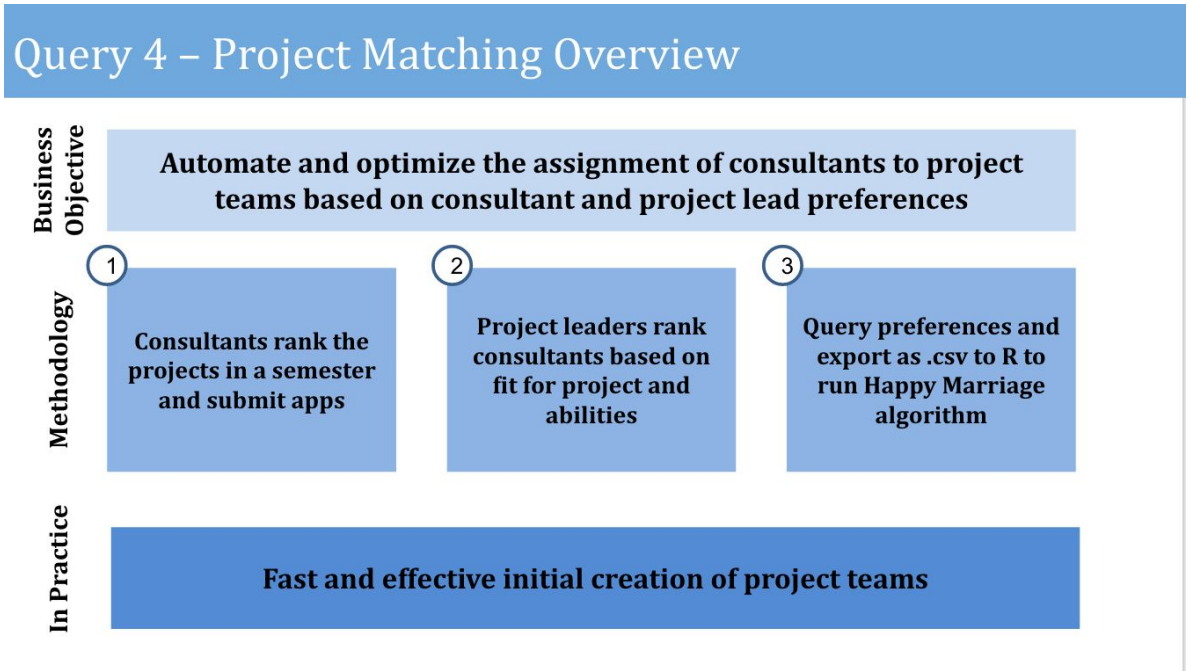
	Justin	Matt	Nicole	Achilleas	Joao	Taylor	Lauren
Justin	0	1	4	12	8	9	13
Matt	1	0	3	7	2	13	9
Nicole	4	3	0	5	8	7	10
Achilleas	12	7	5	0	6	12	5
Joao	8	2	8	6	0	11	14
Taylor	9	13	7	12	11	0	15
Lauren	13	9	10	5	14	15	0

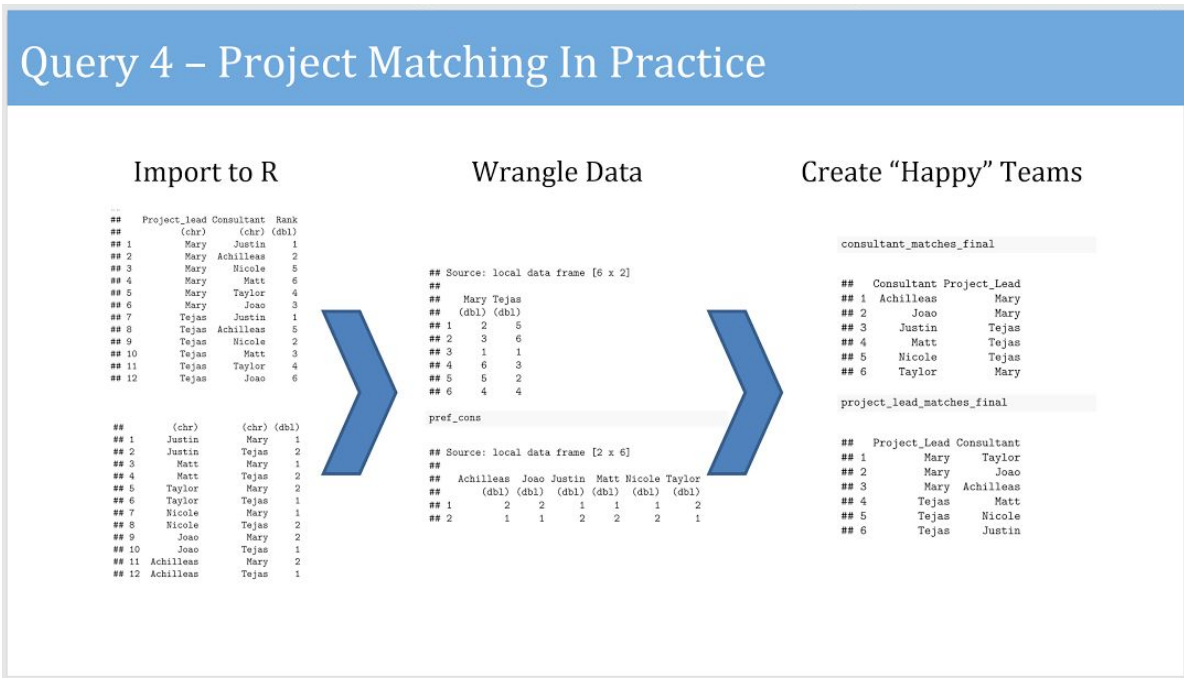
Using the UPGMA hierarchical cluster algorithm, we identify and rank the closeness of relationships using an iterative process. With this analysis, we can produce a dendrogram showing the closeness of relationships among this sample of 7 members from a particular industry so that a member can understand where they are most likely able to make a meaningful connection. The output from the sample above gives us the following dendrogram:



The client showed great enthusiasm for this query as TBG's alumni network is a great resource for members in the club, but identifying people to talk to is very challenging for newer members, especially those who were not in the club at the same time as the current alumni were. This would help new members instantly gain accessibility to TBG's alumni network.

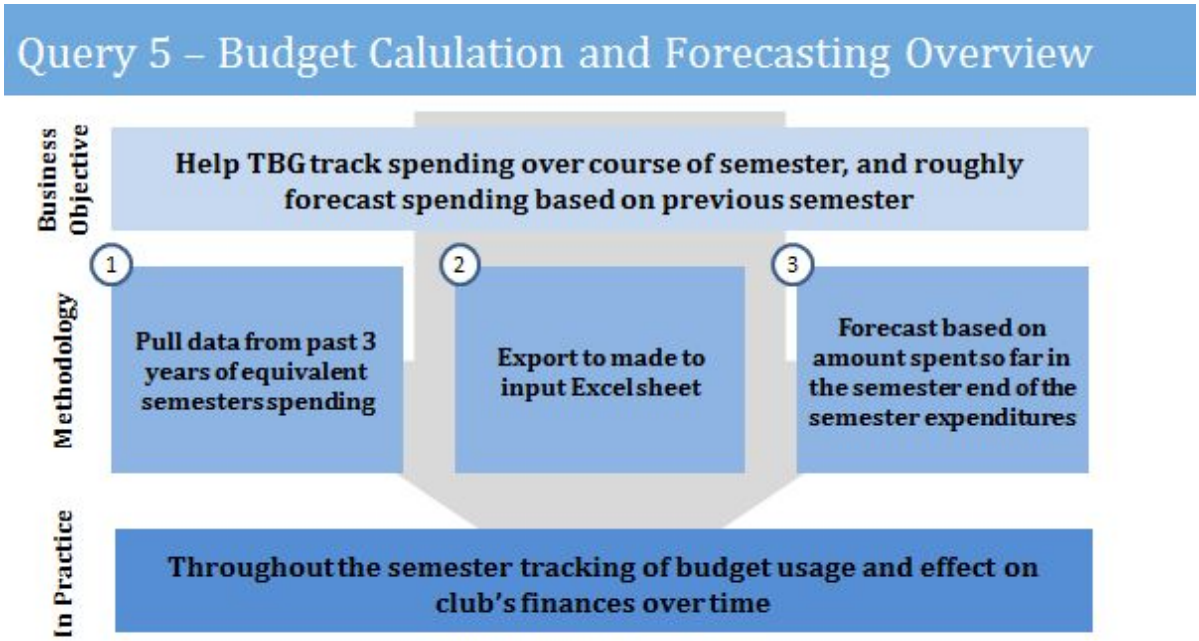
Query 4





The client expressed great enthusiasm for this query. Although they would not blindly accept the teams formed from the algorithm they believe it would provide a better starting point for the discussion and trades that are used to finalize teams. The current process of drafting teams before this discussion ensues takes a couple hours, and the teams often require significant reshaping afterwards, but the stable marriage formed teams would likely be better, and require less edits.

Query 5



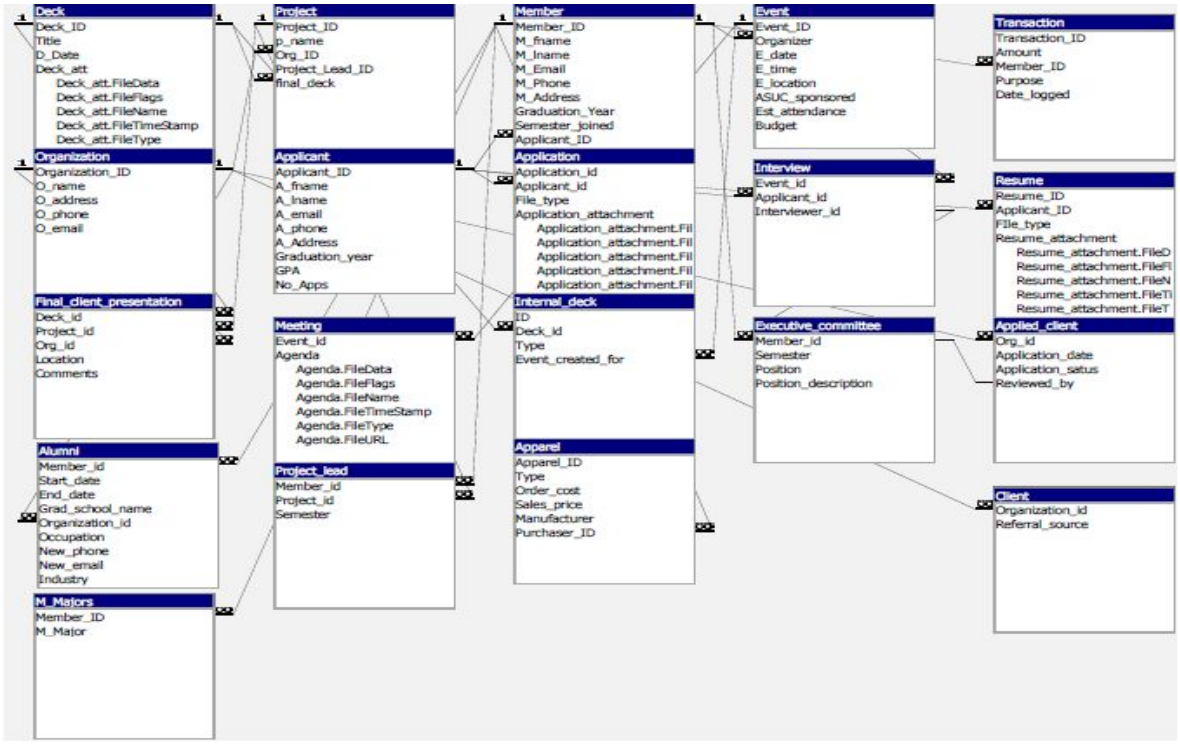
Filtering transactions to expenditures from different years, semesters, and months a forecast can be performed for expenditures for the coming months accounting for semesterly and monthly trends as well as the general increase in costs over time, which TBG does not do well at the moment, and has led to several occasions where allotted funds were insufficient to fully cover expenses (i.e. Insufficient food budget for retreat this past year) A modified Holt-Winter model would be effective here treating the month and semester as seasonal factors:

$$
\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + s^{(1)}_{t-m_1} + s^{(2)}_{t-m_2} + \varepsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \beta\varepsilon_t \\
s^{(1)}_t &= s^{(1)}_{t-m_1} + \gamma_{d_1}\varepsilon_t \\
s^{(2)}_t &= s^{(2)}_{t-m_2} + \gamma_{d_2}\varepsilon_t
\end{aligned}
$$

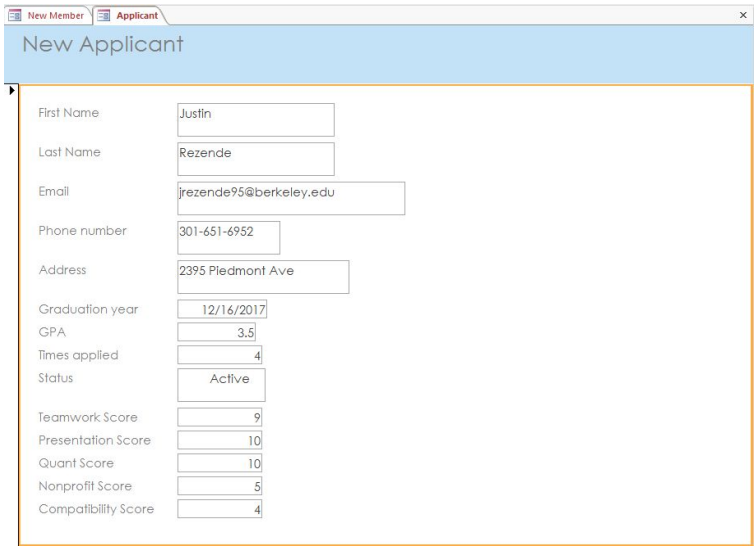
A confidence interval could also be constructed from this method examining the sample variance of the error vectors (sy), to set-up an interval based on a student's t-test with yi ± sy(1-alpha/2). The client also liked this query quite a bit. Although they were a little confused about the nature of the forecasting model they acknowledged their current approach towards budgeting needs improvement.

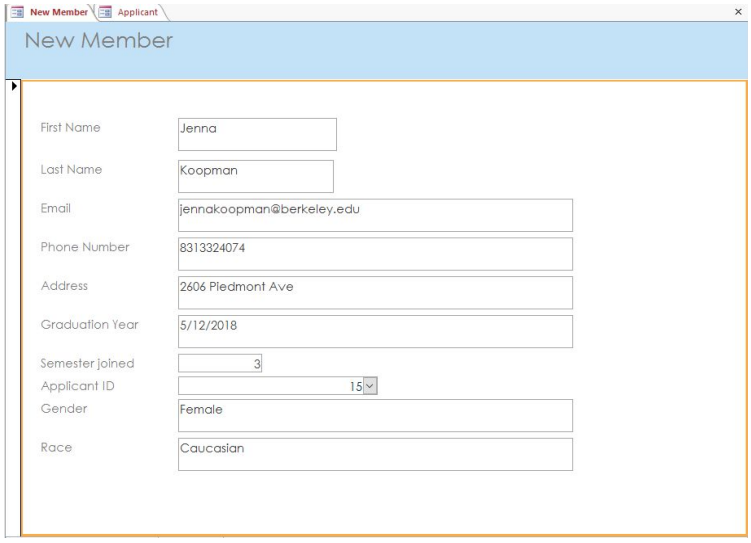
MS Access Model

Abbreviated Relational Database



Forms





Reports

Member Roster

Member ID	First Name	Last Name	Email
2	Justin	Rezende	jrezende95@gmail.com
3	Matt	Robertson	mroberts@berkeley.edu
4	Nicole	Berun	nberun@berkeley.edu
5	Achilleas	Gkinis	agkinis@berkeley.edu
6	Joao	Drummond	joaogabrielfmd@berkeley.edu
7	Taylor	Lyberger	tlyberger@berkeley.edu
8	Lauren	Hanlon	laurenhanlon@berkeley.edu
9	Olivia	Dibenedetto	oliviad@berkeley.edu
10	Ryan	Chapman	ryanchapman@berkeley.edu
11	James	Yorke	jamesyorke@berkeley.edu
12	Patricia	Flynn	patriciaflynn@berkeley.edu
13	Kara	Enomoto	kara@berkeley.edu
14	Danielle	Jarvis	daniellejarvis@berkeley.edu
15	Jenna	Koopman	jennakoopman@berkeley.edu

Transactions

Transaction ID	Member_ID	Purpose	Amount
1	2	membership dues	20
2	2	meeting - food	-10
3	3	donation	200
4	4	apparel cost	-50
5	5	travel expense	-100
6	3	membership dues	50
7	7	social	-150
8	6	ASUC	150
9	3	apparel payments	500
10	4	apparel costs	-500
11	5	event	-50
12	8	flyers	-100
13	8	ASUC	100
14	9	membership dues	275
15	11	social	-160
16	10	apparel payment	230
17	10	apparel costs	-230
		Grand Total	175

Normalization

Relation 1: App_Generates(Transaction_ID, Apparel_ID, type, amount)
Status: 1NF
FDs: Transaction_ID → amount
Apparel_ID → type
Violation: 2NF (Both FDs are determined by partial keys of the relation)
How to normalize further:

- App_Generates(Transaction_ID, Apparel_ID)
- Transaction(Transaction_ID, amount, Member_ID, purpose, date_logged)
- Apparel(Apparel_ID, type, order_cost, sale_price, manufacturer, Purchaser_ID)

Relation 2: Leads(Member_ID, Event_ID, Date, Training_deck)
Status: 1NF
FDs: {Member_ID, Event_ID} → Training_deck
Event_ID → Date
Violation: 2NF (Second FD is partial, date determined by partial key Event_ID)
How to normalize further:

- Leads_event(Member_ID, Event_ID, Training_deck)
- Events(Event_ID, Organizer, E_date, E_time, E_location, ASUC_sponsored, Est_attendance, Budget, Advertising_avenues)

Relation 3: Donates(Organization_ID, Transaction_ID, type, amount, date, semester)
Status: 2NF
FDs: {Organization_ID, Transaction_ID} → {Type, Amount, Date}
Date → Semester
Violation: 3NF (Non-prime attribute determines non-prime attribute in second FD)
How to normalize further:

- Donates(Organization_ID, Transaction_ID, type, amount, date, semester)
- Date_translate(Date, Semester)

Relation 4: Alumni(Member_ID, Start_date, End_date, Grad_school_name, Organization_name, Occupation, New_phone, New_email, Industry)
Status: 2NF
FDs: {Member_ID, Start_date} → {End_date, Grad_school_name, Organization_name, Occupation, New_phone, New_email, Industry}
Organization_name → Industry
Violation: 3NF (Non-prime attribute determines non-prime attribute in second FD)
How to normalize further:

- Alumni(Member_ID, Start_date, End_date, Grad_school_name, Organization_name, Occupation, New_phone, New_email, Industry)
- Org_Industry(Organization_name, Industry)

Relation 5: Project_advisor(Member_ID, Semester, Project_ID)
Status: 3NF
FDs: Member_ID → Project_ID
Project_ID → Semester
Violation: BCNF (FDs are not determined by a superkey)
How to normalize further:

- Project_advisor(Member_ID, Project_ID)
- Project(Project_ID, Semester, p_name, Organization_ID, Project_Lead_ID, final_deck)

Client Feedback

TBG was represented by Justin Rezende at the presentation, though he has had conversations throughout the semester with various leaders in the organization to keep the project focused on maximizing its value add to TBG. They were most enthused about the ability to organize all the data in one central location as the organization is frustrating to many in the club. The client especially liked the Alumni Matching, Project Matching, and Marketing Strategy Success Queries as the club is striving to improve alumni accessibility, streamline internal processes, and validate the outreach efforts of its members.

Team Member Contributions

Taylor Lyberger: COO. Contributed heavily to the relational schema organization, presentation, and normalization. Created cluster analysis sample for query 3. Kept track of and focused on incorporating Professor Goldberg's feedback.

Justin Rezende: CEO. Developed ideas for several queries, wrote SQL for a couple, and performed research on methodologies to utilize pulled data. Kept project focused on adding client value, organized meetings, and created first and final presentation slide decks.

Lauren Hanlon: Main contribution was with query creation and SQL. Helped develop ideas for several queries, then wrote most of SQL code to pull data. Helped with putting together presentations, initial EER diagram and revising queries from feedback in DP II/III.

Nicole Benun: Worked on initial EER diagram relationships, helped develop several query ideas, identified normalization needs and solutions in the relational schema, contributed to each presentation organization and delivery. Ensured team deliverables were prepared by the due date.

Joao Drummond: Contributed to EER diagram design, setting cardinality constraints and clearing up the organization. Worked on the implementation of the relational schema by identifying relationship tables and helped in finding relations that required normalization.

Achilleas Ghinis: Contributed to the relational schema design and implementing our MS Access database functionality, including creating and linking tables and running queries, worked on the implementation of the project matching algorithm in R. .

Emily Rah: Organized EER diagram clearly, adding to and fact-checking the cardinality constraints and ensuring there were no dangling subclasses. Aided in relational schema creation as well as MS Access database organization by creating the first group of tables and adding to them later on.

Matt Campbell: CCO. Contributed to query formulation and analysis throughout iterations of the design project. In addition he participated in presentation creation and delivery.

Discussion and Reflection

Our group is very pleased with our database, queries, and final deliverables. We feel that we were able to really cater to the data needs of TBG, or a similar style consulting group or organization. In the idea formation stage, our group came together quite effectively to brainstorm the framework for our database. We brought in several members of TBG and listened to their current record and data management and what their biggest needs were that were not being met.

Once we had a good understanding of our client we iterated through several EER diagrams and relations schema, each time taking feedback from the GSI, professor, and members of TBG, and applied the new concepts and lessons we learned every week, finally ending in our functional Access database. For our query formation we collectively sorted through dozens of ideas and fleshed out the ones that would not only serve our client the most in the present, but would also demonstrate our creativity and show the great advantages that an advanced database could bring our client.

However, one important area of improvement would be in how we handled the balance of meeting the data needs of our client versus the potential needs or features of a fully fledged organization. Obviously, TBG's immediate needs do not necessarily warrant the most advanced database as they are a small, but growing, organization. When talking with members about the project, many of our proposed ideas were met with reservations. For example, one of our queries had to do with candidate scoring and predicting how likely a certain person would be accepted based on past data. Obviously there was not a lot of interest from TBG, as they favor an exclusively subjective process that is not necessarily scalable. While we still kept this query, we believe that we could have pushed our client a little further into showing the possibilities of a database at scale, rather than for their current place.

Our future work will mainly spin off of this problem. We can really benefit and add value to an organization like TBG by showing them how they can advance their organization with a robust database. We would like to build out more functionality regarding alumni and active matching through common interests, job fields, or mutual contacts. One of the most powerful advantages of a group like TBG is there robust alumni network, and we would like to help strengthen that. Another topic that was discussed for future work was is in regards to fundraising. We could help store donor and corporate connections and leverage that data to maximize the funding that TBG and their clients receive. TBG is really a special organization, whose mission it is to empower non-profits across the Bay Area so that they can help as many lives and causes as possible. It is our hope that showing them the power an intelligently designed database will further empower the nonprofits they serve.